

## Patent Abstracts of Japan

NCA

PF030127AB+AC

CITED BY APPLICANT

PUBLICATION NUMBER : 07044575  
PUBLICATION DATE : 14-02-95

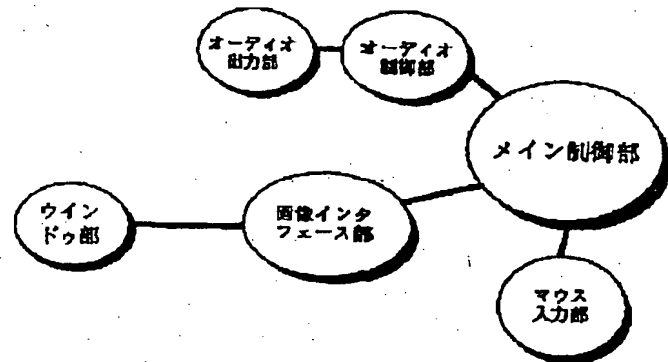
APPLICATION DATE : 03-08-93  
APPLICATION NUMBER : 05192447

APPLICANT : OKADA KENICHI;

INVENTOR : OKI NAOTO;

INT.CL. : G06F 17/30 G09F 27/00 G11B 27/10  
G11B 27/34

TITLE : VOICE INFORMATION RETRIEVAL  
SYSTEM AND ITS DEVICE



ABSTRACT : PURPOSE: To retrieve a desired sound source from among the plural sound sources, and to attain the data display by a retrieval interface equipped with a virtual sound field space and a visualizing means corresponding to the virtual sound field.

CONSTITUTION: In a voice information retrieval system, an access to the desired sound source is performed from among the plural sound sources based on the direction and distance of the desired sound source by the retrieval interface equipped with the virtual sound field space and the visualizing means corresponding to the plural sound sources in the sound field space. Then, the visualizing means appears at each prescribed point by moving in a video and the sound field space with a device such as a mouse, and at the time of reaching the desired sound source, the video of the sound source and the characteristic are outputted. A mouse inputting part which inputs the instruction of a user, picture interface part which stores and offers a picture, and audio control part which controls pronunciation are connected with a main control part which controls the entire system, and an audio outputting part is connected with the audio control part, and a picture outputting part is connected with the picture interface part.

COPYRIGHT: (C)1995,JPO

JPO7044575.pdf

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-44575

(43) 公開日 平成7年(1995)2月14日

(51) Int.Cl. <sup>4</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30				
G 0 9 F 27/00	N	7323-5G		
G 1 1 B 27/10	E	8224-5D		
27/34	N	8224-5D		
		9194-5L		
			G 0 6 F 15/ 40	3 7 0 E
			審査請求	未請求 請求項の数 5 O L (全 9 頁)

(21) 出願番号 特願平5-192447

(22) 出願日 平成5年(1993)8月3日

(71) 出願人 391023987

松下 温  
東京都新宿区喜久井町36

(71) 出願人 392008231

岡田 謙一  
東京都文京区本郷4-25-12

(72) 発明者 松下 温

東京都新宿区喜久井町36

(72) 発明者 岡田 謙一

東京都文京区本郷4-25-12

(72) 発明者 大木 直人

神奈川県横浜市港北区日吉三丁目14番1号  
慶應義塾大学理工学部計測工学科内

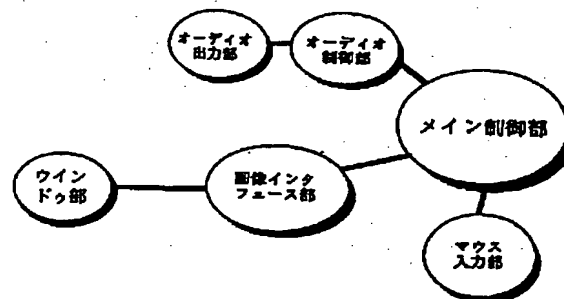
(74) 代理人 弁理士 鈴木 正次

(54) 【発明の名称】 音声情報検索システム及び装置

(57) 【要約】

【目的】 この発明は、仮想的な音場空間と、これに対応する視覚手段を具えた検索インターフェースにより多数の音源中から求める音源を検索し、そのデータ表示を受けることを目的としたものである。

【構成】 想的音場空間と、当該音場空間の多数の音源に対応する視覚手段とを具えた検索インターフェースにより、前記多数の音源中から、求める音源の方向・距離を頼りにこれにアクセスし、マウスなどのデバイスで、当該映像及び音場空間中を移動すれば、所定のポイント毎に当該部の視覚手段を現わし、求める音源に到達した場合に、当該音源の映像とその特性などを出力することを特徴とした音声情報検索システム。システム全体を制御するメイン制御部に、ユーザの指示を入力するマウス入力部と、画像を蓄積及び提供する画像インターフェース部と、発音制御のオーディオ制御部を夫々連結し、前記オーディオ制御部にオーディオ出力部を連結し、画像インターフェース部に画像出力部を連結したことを特徴とする音声情報検索装置。



1

## 【特許請求の範囲】

【請求項1】 仮想的音場空間と、当該音場空間の多数の音源に対応する視覚手段とを具えた検索インターフェースにより、前記多数の音源中から、求める音源の方向・距離を頼りにこれにアクセスし、マウスなどのデバイスで、当該映像及び音場空間中を移動すれば、所定のポイント毎に当該部の視覚手段を現わし、求める音源に到達した場合に、当該音源の映像とその特性などを出力することを特徴とした音声情報検索システム。

【請求項2】 仮想的な音場空間は、人間の聞き分け可能な範囲に音源を配置する空間とすることを特徴とした請求項1記載の音声情報検索システム。

【請求項3】 仮想的な音場空間を多階層化し、アクセス域を絞り込み可能としたことを特徴とする請求項1記載の音声情報検索システム。

【請求項4】 視覚手段は、レーダーウインドウ、鳥瞰ウインド、3Dウインド及びデータ表示ウインドとしたことを特徴とする請求項1記載の音声情報検索システム。

【請求項5】 システム全体を制御するメイン制御部に、ユーザの指示を入力するマウス入力部と、画像を蓄積及び提供する画像インターフェース部と、発音制御のオーディオ制御部を夫々連結し、前記オーディオ制御部にオーディオ出力部を連結し、画像インターフェース部に画像出力部を連結したことを特徴とする音声情報検索装置。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、仮想的な音場空間と、これに対応する視覚手段を具えた検索インターフェースにより多数の音源中から求める音源を検索し、そのデータ表示を受けることを目的とした音声情報検索システム及び装置に関する。

【0002】

【従来の技術】 従来知られているリンク法は、文字データ・数値データを主に扱うためのデータベースで、拡張機能としてマルチメディアデータを扱うときによく用いられる方式であり、最も簡便な方法である。この方法では、マルチメディアデータは1つ、場合によっては複数の文字データ・数値データとリンクされ、ある検索条件によって文字データ・数値データが検索されると、それらとリンクされている画像データ・音声データが同時にユーザに提供される。

【0003】 またインデックス法は、従来のデータベースにおけるインデックス検索をそのままマルチメディア情報に応用したものである。各データの整理番号あるいはデータ名と、そのデータが格納されている記憶デバイスのアドレスとの対応表がインデックスである。前記による場合の検索は、ユーザが整理番号あるいはデータ名を検索システムに入力し、システムが入力された整理番

2

号あるいはデータ名をインデックスと照らしあわせ、そのデータの記憶デバイス上の位置を得て、目的のデータを出力するという手順で行われる。

【0004】 このインデックス法を用いたシステムの身近な例として、CDがある。CDには曲目（データ）に応じてトラック番号があり、ユーザはデータとトラック番号との対照表（インデックス）を見て、聴きたい曲のトラック番号を知ることができる。この番号をCDプレーヤに入力することによって、聴きたい曲を聴くことができる。こうしてみると、CDはそれ自身が立派な音声データベースと言える。

【0005】 次にキーワード法は、前記のインデックス法を使いやすく改良したもので、各データにそのデータの属性など関連する複数のキーワードを付加し、それらのキーワードによって検索しようという方法である。多くのマルチメディアデータベースがこの方法を採用している。

【0006】 更にパターンマッチング法は、認識技術が発達し、コンピュータに対しマルチメディア情報を入力し、それを認識・内容解析することが可能となり、音声のデータを音声の検索条件によって検索することが実現した。このような検索方式は、キーワード検索のようにメディア間の変換（音声データから文字データへ、又は文字データから音声データへ）を経る必要がないため、より適切な音声情報検索が可能である。鳥類図鑑のハイパーブック（Hyperbook）では、検索したい鳥の鳴き声をユーザが真似ることによって検索条件を入力することができる。ユーザの鳴き真似を解析し、振幅構造・ピッチ構造・周波数成分変化構造などの音響的特徴を解析し、登録されている鳥の鳴き声データの音響的特徴との照合を距離関数によって行う。

【0007】

【発明により解決すべき課題】 前記リンク法は、あくまでもマルチメディアデータは文字データ・数値データの付録として扱われ、それ自身を直接検索することが困難という問題点があった。

【0008】 またインデックス法では、ユーザが検索したい情報の整理番号・データ名を知らない場合には検索が非常に困難になる問題点がある。例えばユーザがCDの曲名と曲番号との対応を調べないと（覚えていないと）聴きたい曲を容易に検索できないことが往々ある。

【0009】 次にキーワード法では、音声データや画像データに含まれる情報を完全にキーワードで表現することは不可能である。例えば「ドビュッシー」の曲「月の光」やキリコの絵、「町の神秘と憂愁」が持つ雰囲気や完全に言葉で表現しようとしても無理なように、どんなに熟慮されたキーワードでも、結局はそのデータを聴いてみるか観てみるかしない限りは、その検索結果が適当か否か判断できない。

【0010】 更にパターンマッチング法の問題点は、音

3

声による検索条件の入力である。例えば鳥類図鑑で鳴き真似を採用することによってこの問題の解決を試みているが、鳥の鳴き声を事細かに表現するのは困難である。

【0011】

【課題を解決するための手段】然るにこの発明は、前記従来方式の各問題点を解消する為に人工現実感を用いた新たな検索方式を提供するインターフェースを提案するものである。前記新たな方式は、インタラクティブ (interactive) な音場インターフェースでは、ユーザがマウスなどのデバイスで自分の動きをインターフェースに伝え、インターフェースは、その動きに応じた音場をユーザに提供し、これによってユーザは恰も仮想的な空間で自分が移動したかのように感じることができる。さらに音場の動きに応じて次の行動を起す。このようにユーザはインターフェースと情報を交換しながら、仮想的な空間を動き回ることができるので、これをインタラクティブな音場インターフェース (interface with interactive sound field、以下ISFという) と名づけた。

【0012】ISFは、音声データベースに登録されている音声データを仮想的な音場に配置して、ユーザはそれらの音声データが発する音の方向、距離を手掛かりに音場空間を移動し、欲する情報に辿り着く検索インターフェースである。ある情報に近づくことによって、そのデータの持つ音声情報だけでなく、画像データなど他のメディアの情報をディスプレイ上に表示することもできる。

【0013】即ちこの発明は仮想的な音場空間と、当該音場空間の多数の音源に対応する視覚手段とを具えた検索インターフェースにより、前記多数の音源中から、求める音源の方向・距離を頼りにこれにアクセスし、マウスなどのデバイスで、当該映像及び音場空間中を移動すれば、所定のポイント毎に当該部の視覚手段を現わし、求める音源に到達した場合に、当該音源の映像とその特性を出力することを特徴とした音声情報検索システムである。

【0014】また他の発明は仮想的な音場空間は、人間の聞き分け可能な範囲に音源を配置する空間とすることを特徴としたものであり、仮想的音場空間を多階層化し、アクセス域を絞り込み可能としたものである。

【0015】次に他の発明は視覚手段として、レーダーウインドウ、鳥瞰ウインド、3Dウインド及びデータ表示ウインドを採用したものである。

【0016】更に装置の発明は、システム全体を制御するメイン制御部に、ユーザの指示を入力するマウス入力部と、画像を蓄積及び提供する画像インターフェース部と、発音制御のオーディオ制御部を夫々連結し、前記オーディオ制御部にオーディオ出力部を連結し、画像インターフェース部に画像出力部を連結したことを特徴とす

4

る音声情報検索装置である。

【0017】ISFは音声情報検索インターフェースとして以下の特徴を持っている。

1. 検索条件の入力を必要としない。

(1) 明確な検索条件が不要。

(2) 検索条件のメディア変換が不要。

(3) ブラウジングによる情報検索。

(4) 簡潔なインターフェース。

2. 音場が左右の定位感、距離感を持つ。

3. 複数の音声データを同時に聴くことができる。

4. データ空間をアプリケーションに応じて演出することが可能である。

5. 音声情報の階層化が可能である。

【0018】前記特質について更に詳細に説明する。

【0019】前記したように、検索条件を入力しなくても情報が得られるという長所は、いくつかのメリットを派生する。例えば検索条件が不要なのであるから、従来のデータベース検索のように確固とした検索条件が無くても、“確かこんな音だった”という漠然としたイメージがあれば検索が可能である。

【0020】また、音を聴きながら情報を検索するのであるから、キーワード検索のように検索条件を無理遣り他のメディアに変換すること無く音声情報の検索ができる。

【0021】ユーザが本をバラバラめくりながら情報を探するときのように、あれこれと情報を“つまみ食い”しながら欲しい情報を探すという方法は、ブラウジングと呼ばれており、画像インターフェースの分野ではよく知られている。もしかすると、これらブラウジングの途中で見られる情報の中にも有用な情報があるかも知れないし、現実世界では、そのような偶然チラッと見た情報が、実は有用な情報であるというケースは、研究のための文献を探しているときに誰もがよく経験することである。このような偶然の情報の発見は、従来のデータベース検索においては不可能である。このようにブラウジングは、人間的なインターフェースとして有効であり、これを音声情報にも適用できるようにしたのである。

【0022】また、従来のデータベースシステムでは、検索のためのインターフェースが非常に複雑であった。これは検索条件を入力する作業が煩雑である場合が多い。検索条件の設定はどうしても文字を入力する作業が必要になるし、全ての操作をマウスで行えるようなシステムも提案されているが、やはり検索条件を設定する作業が必要であることには変わり無い。この点ISFでは、ユーザは自分が音空間の中でどこに移動したいかを、入力デバイスを用いてシステムに指示するだけでよい。現在のところ入力デバイスはマウスを用いているが、方向が指示できさえすればよいので、キーボードの10キーでもデータグローブでも操作は可能である。

【0023】ISFの音インターフェースは、左右の定

位感・遠近感を伴って提供される。このような立体感を伴った音場は、ユーザがより自然な形で音声データを聴取できる環境を実現し、これによって同時に発せられる複数の音声データを聞き分けることが可能となる。

【0024】人間には、同時に複数の音声が存在する環境で、ある特定の音声だけを聴き取る能力がある。例えば、人が大勢集まってガヤガヤと会話が飛び交っているカクテルパーティのような場所でも、自分の名前が会話に出てくると、パッと振り向くことができる。これになぞられて、このような効果はカクテルパーティ効果と呼ばれる。カクテルパーティ効果は、個々の音源が独自の左右定位・距離定位を持った音像である場合により効果が強く現れる。そこで、音声データベースに登録されているいくつかのデータを個々の左右定位・距離定位で再生する事により、それらを識別・比較することができる。

【0025】仮想的な音空間に音声データを配置する際に、ただ無秩序にデータを配置したのではユーザに混乱を招くだけである。そこで、仮想的な音空間を何かに似せて演出することを考えた。

【0026】ISFでは様々な演出が可能だが、この発明で例示システムとした“ISF昆虫図鑑”では、自然界を模しており、この空間には草原があり、木のざわめく林があり、せせらぎの聴こえる川が流れている。このような音で構成された仮想的な空間に、“鈴虫は草原に”“ミンミンゼミは林に”という具合にソース（音声データ）が配置されている。このためISFにおける仮想的な音空間は、ソース（音声データ=虫など）とオブジェクト（音声データ以外のもので、音を出すものもある=木、川など）で構成されている。ユーザはこのような仮想的な音空間の中をあたかも自分がその中にいるような感覚で歩き回り、情報を得ることができる。

【0027】人間にカクテルパーティ効果があるとはいえず、百や千もの音を同時に聞いて、その個々の音を聞き分けることは不可能である。一方、ハードウェアの面でも制約がある。そこで今回実装した試作システムも同時に8種類の音を独立した音声ラインに送出するのが限界である。このため、人間の側からも、ハードウェアの側からも同時に発生される音の種類は限られてくる。そこで、実用化を進める為には音声情報を階層化することが好ましい。

【0028】例えば音声情報をあらかじめいくつかグループ分けしておき、あるグループに属する音声情報は仮想的な音空間内のあるまとまった範囲に位置する。ユーザがあるグループの外ににいるときには、そのグループの代表音がグループの重心に音像定位して聴こえている。グループの代表音は、ある一つの音声情報でもよいし、グループ全体のまとまった音でもよい。ユーザがグループの中に入ったときのみ、グループ内の個々の音が

それぞれの音像定位を持って音を発する。これによってより多くの音声情報が取り扱えるようになる。

【0029】例えば、昆虫のなかで、蟬というグループを作っておき、ユーザがグループの外ににいるときには蟬時雨がグループの中心辺りから聴こえるが、グループに入り込むと個々の蟬がそれぞれの場所で鳴いているのが聞こえてくるといった具合である。

【0030】前記ISFの実装には、エフェクタによる方法を用いることができる。この方法によれば、疑似的に音の遠近感を表現するので、これにより、ウインドシステム上で、ウインド、アイコンの前後関係を遠近感として表現する。例えば遠方を表現するためにリバーヴ（残響）の効果を強くしたり、背後からの音を表現するためにロウパスフィルタなどを用いる。この方法によれば、左右方向の音像定位は、左右の音量差の他に、左右の信号の時間差によっても得ることができる。

【0031】前記エフェクタ法を用いれば、MIDI規格のインターフェースによって、コンピュータと相互に情報のやり取りが可能であり、制御しやすい長所がある。

【0032】然し乍ら、この発明のISFの実装に際し、前記エフェクタ法以外にも、右信号と左信号の相関係数を1から-1まで連続的に変化させる相関係数変化法を採用することができる。

【0033】またバイノーラル方式によれば人間の頭部を音響的に模したダミーヘッドの両耳の鼓膜に相当する部分に設置された2個のマイクロフォンによって録音される。ユーザはこれをヘッドフォンによって聴くことにより、ダミーヘッドがおかれた位置に居るかのような音像の定位感を得ることができるバイノーラル方式を採用することができる。尤も現在バイノーラル方式はダミーヘッドを用いずにダミーヘッドの音響的性質の伝達関数を測定し、それを表わすデジタルフィルタを用いて、バイノーラル効果を得ている。この方式によれば、前後方向と左右方向の音像定位をまとめて得られる特質がある。

【0034】前記各実装方式には夫々得失があるので、この発明の実施に際し、対象物の特性と合致した方式を採用する。例えば高周波成分を多く含んでいる昆虫検索にはエフェクタ法による実装が好ましい。

【0035】またバイノーラル方式では、デジタルフィルタで行われる畳み込み積分に必要な計算時間が非常に長く、システム全体としてリアルタイムな反応が期待できないという問題点があるが、この点を解消できれば十分採用し得る方式である。

【0036】要するにこの発明は実装方式に限定を受けないものであり、対象音源に適する方式を適宜採用することができる。

【0037】

【作用】この発明によれば、多数の音の中から求める音

に絞り込むのであるから、予め検索数値その他の記号等の入力とは全く不用となる。

【0038】また仮想的な音場空間と視覚手段の組み合わせによってユーザの臨場感を向上し、求める音の選定を適格にすることができる。

【0039】次に複数の視覚手段を併用することにより、求める音源への移動を適確にし、絞り込みの迅速化を図ることができる。

【0040】

【実施例】この発明の実施例を図1、2に基づいて説明する。

【0041】図1は、この発明のハードウェアの構成を示すものである。即ちメイン制御部と、マウス入力部、画像インターフェース部、オーディオ制御部を連結し、オーディオ制御部には、オーディオ出力部を連結し、画像インターフェース部には画像出力部してある。

【0042】図2は、前記図1の構成に基づく機器を示すものである。即ちスパークステーション (spark station) は、RS-232Cを介してパソコンのNEC PC-9801に接続し、PC-9801はミディ (MIDI) を介し、サンプラー (sampler) EPS-16に接続し、サンプラーEPS-16は、ミディ、オーディオ (audio) を介してミディミキサーDMP-11に接続し、ミディミキサーDMP-11はオーディオを介してステレオヘッドホン (stereo headphone) に接続している。

【0043】前記スパークステーション1は、システム全体を制御する装置である。そこでディスプレイにより視覚情報を提供し、マウスによってユーザは音場空間を移動する。

【0044】前記DP/4は、4チャンネルの独立した音声信号にそれぞれ独立した処理を行うことができる。これらの処理はMIDI信号によるリアルタイム制御ができる (図2)。相関係数変化法を用いて実装しているので、音像の距離感制御のために、音声信号の位相制御を行う。

【0045】また前記以外の音声信号の処理として、距離感を提供するリバーブ (残響) ・ディレイ (遅延) 、音像の広がり感を提供するフェイズシフト・コーラス、音質を変させるイコライザなどが可能である。

【0046】またEPS-16はソース・オブジェクトの音源であって、MIDI信号により制御ができる。1MBのメモリに最大32種類の音をデジタル録音し、16種類の音を同時に再生し、8チャンネルの独立したラインから音声信号を出すことができる。このEPS-16の音声データのサンプリングレイトは最高44、8kHzであり、CDと同等の音質を得ることができる。

【0047】前記EPS-16、DP/4によって得られる複数の音像を、1つの音場にまとめるためにDMP-11を用いる。DMP-11は、MIDI信号での制

御が可能な8チャンネルデジタルミキサーである。8チャンネルのアナログ音声ラインを入力時にA/D変換し、その後の様々な処理はDSPを用いてデジタル信号のまま行なう。このため音質の劣化とか、ノイズの発生を抑えることができると共に、各チャンネル独立にパンニング (左右音像定位) ・リバーブ (残響効果) ・イコライジング (周波数成分変更) 等もできる (図2)。

【0048】前記のようなMIDI機器を用いずに、ワークステーションがこれらの機能をサポートしているのが望ましいが、現開発の実情としては一部の機能が提供されているのみである。例えばサンブラの代りにワークステーションの外部記憶装置に音声データを蓄えることもできるが、音声データを2チャンネル以上の複数の独立したラインから出力することはできない。今後マルチメディアの発展により、これらの機能をワークステーション機能の一部として組み込まれた際には、この発明にも採用することができる。

【0049】前記実施例の音像制御における音像定位に関するパラメータは以下の4種である。

- 【0050】1. 音量はユーザとソース・オブジェクト間の距離の自乗に反比例する。
2. 左右定位はパンを用いて360°、32方向に音像を振り分ける。
3. 距離感感度はリバーブを用いて、距離が遠いほど残響音のレベルを大きくしている。
4. 前後はユーザの背後から聞えるべき音を、ローパスフィルタを用いてこもった音にしている。

【0051】前記距離感に関し、実装に相関係数変化法を用いた場合には、ホワイトノイズについては比較的良好的な遠近感が得られたが、この実施例のように、虫の鳴き声を検索する場合には良好な距離感が得られなかった。この実施例ではエフェクタによる方法を用いた。前記のように、虫の鳴き声を検索する場合に、相関係数変化法が不利な理由は次のように考えられる。

【0052】即ち相関係数変化法は、前後方向の移動音像に対して距離感が顕著に現れること、周波数によって距離感が異なり、高い周波数域では距離感が乏しいこと、及び放送などで実際に用いられる時には、他の距離感を出す方法を組み合わせて用いられているなどである。

【0053】従って音源の性質によっては、この発明の実施に十分採用できる方法である。

【0054】前記DMP11・EPS16を制御するMIDIデータは、個々が6バイトの情報であり、図3に示す構造をもっている。即ち第1、2バイトは、制御するMIDI機器に割り当てられたチャンネル番号である。例えば、DP/4が1ch、EPS-16plusが2ch、DMP-11が3chと4chにチャンネル番号が割り当てられている。第2、3バイトは制御すべき機能の番号であり、第4、5バイトがその機能のパラメータである。これらの値はMIDI機器側で決めら

9

れており、それを用いている。

【0055】前記MIDIデータ生成部では、音像位置計算部で計算された音像定位のための各パラメータを、前記データフォーマットでMIDI信号に変換する。前記MIDIデータ生成部で作られたMIDIデータを、スパークステーションのRS-232Cポートから出力する。現バージョンではハンドシェイクなどのエラー防止装置は行っていないので、この通信エラーの対策はMIDI信号中継部で行うようにしてある。この発明においては、音のインターフェースだけでなく、音源の方向、距離感を判断すると共に、臨場感を表現する補助的

10 インターフェースとして画像を用いたインターフェースをいくつか採用しているの、その実施例について説明する。この画像インターフェースは、X-ウィンド上に、X-ビューを用いて実現されている。図4はコントロールウィンドの一例である。

【0056】このコントロールウィンドは、この発明のシステム起動時の初期画面で存在する唯一のウィンドウであり、ユーザのマウス操作は主としてこのウィンドウで行われ、マウスインターフェース部もこのコントロールウィンドウに属する。

【0057】このウィンドウ内には、各種画像のウィンドウを開くボタン(a~c)と、システムを終了するボタン(d)及びマウスの動きを感知するキャンパス(e)がある。

【0058】ユーザは前記キャンパス(e)内でマウスを動かすことにより、その方向の音場空間を進むことができる。またマウス上の左右のボタンにより、ユーザは

30  $360/16=22.5$ 度ずつ向きを変えることができる。即ち図5において、ユーザは16方向の内、任意の方向を選択して音場空間を進むことができる。

【0059】マウスインターフェース部では、この動きによりユーザの位置を変更し、それぞれのウィンドウ及びオーディオ制御部に情報を伝達する。またマウスインターフェース部では、コントロールウィンドウ内のマウスの位置の制御を行っている。マウスが1度キャンパスe(図4)の中に入り、ユーザが音場空間内を動かそうとしてマウスがキャンパスから出ようとする、マウスの位置をキャンパスの中央に戻す制御をする。もちろんマウスが戻されたとしても、この時ユーザの位置は変更

40 されない。つまりユーザはコントロールウィンドウ上のマウスの位置を気にせず音場空間をさまようことができる。これによりユーザは目を閉じたまま音だけを頼りに動きまわることもできるし、開かれている他のウィンドウに集中して動きまわることもできる。

【0060】この実施例におけるレーダーウィンドウは、ソースを仮想空間の上方から見た図(図6)を表示するウィンドウであり、オブジェクトは表示しない。ソースとユーザとの相対的な位置が判るので、ソースを目

ざしてマウスを動かし易い特質がある。

10

【0061】図6に示すように、ユーザはレーダー画面上で白い点により表示され、ソースの種類による区別はされていない。またユーザの前方に相当する方向は縦線の上方向(画面上の上)であり、マウスのボタンのクリックによりユーザの方向が変わったとしても変化しない。例えばユーザがマウスの左ボタンをクリックして左に旋回したとすると、レーダーウィンドウは画面上のソースをユーザを中心として右に回転して、ユーザが常に前方(レーダー画面の上方)を向いているように保っている。レーダーウィンドウ上にはセンシティビティ(SENSITIVITY)というスライダーバー(図6b)がついており、これはレーダーの感度を変えるものである。この感度を変えることによりレーダーに映る範囲が変化し、その値がスライダーバーの横及びレーダー画面の上方に表示される。

【0062】センシティビティの値は大きい方がレーダーに映る範囲が広くなり、小さい方が狭くなっている。この感度は任意に好まないときに変えることができるので、ソースが近くにならないときには感度を大きくし、目的のソースに近寄ってきたら感度を小さくしてユーザの位置の微調整するというような使い分けができるようにしてある。また視覚的感度を判別できるようにレーダー画面の縦線と横線のダッシュパターンが感度値により変化するようにしてある。

【0063】この実施例における鳥瞰ウィンドウ(図7)は、レーダーウィンドウとは表裏の関係になっており、ソースを表示せずにオブジェクトのみを表示させ、ユーザ周辺の風景の上からみた図を表示するものである。

30 【0064】このウィンドウでは、ユーザはソースがありそうな場所(例えば蟬ならば森の中)をソースの音をたよりに動きまわることになる。

【0065】このウィンドウもレーダーウィンドウと同様に画面の中心にユーザが位置し、画面の上方向がユーザの前方と一致している。鳥瞰ウィンドウは、例えば50×50のマップデータに対応したマップの中から、ユーザの周囲、前後左右5個のオブジェクトに相当する範囲がユーザの見える範囲となっている。またユーザのマウスの動きによりスクロール、ボタンのクリックにより回転するようになっている。

40 【0066】実施上は、例えば図8のように、キャンパス上に16方向それぞれに対応した50×50の16個のマップが画いてあり、鳥瞰ウィンドウにはその一部をビューとして表示している。そしてユーザがマウスのボタンをクリックしたという情報が入ってくると、その方向に相当したマップにジャンプするようになっている。これによりユーザの前方は常に変えることはない。このように予め16種類のマップを用意しておくことは、メモリ効率上不利になるが、そのつど書き換えることによる

50 時間を削減することができる。またこのときに生じる画

面のちらつきもなくなる。この方法によれば、リアルタイムをとることができる。前記における鳥瞰ウインドウに使用するビットマップは、アイコンエディタにより作成されており、例えば図9の通りである。実際にユーザに見えるそれぞれのオブジェクトの大きさは $16 \times 16 = 256$ ドットであるが、ビットマップは正方形のため $16 \times 16$ のものをを使うとユーザの向いている方向90度の倍数以外のときにはマップ上に何も無い空間が存在してしまう。そこで $24 \times 24$ の大きさのビットマップを使い、重なる部分はそれぞれのオブジェクトの論理和をとることにより、前記の空間の存在を排除している。

【0067】この実施例における3Dウインドウ(3次元オブジェクトのウインドウ)(図10)は、鳥瞰ウインドウと同様にユーザ周辺の風景だけを表示するものであるが、ここでは3次元的にオブジェクトを配置している。これは風景をユーザの視線に近付け、臨場感を高めるためである。またこのウインドウでは、ユーザの方向感覚認識のための背景も表示している。ウインドウ画面に入る背景はユーザの視野角の90度に相当する範囲としてあり、この視野に入るオブジェクトを画面に表示している。ユーザのとり得る方向は16通りあるので、背景も16通り用意してあり、マウスを1回クリックすることにより背景は画面の4分の1ずつ変化していくようになっていく。それぞれのオブジェクトはユーザとの距離に従い、4~6種類用意してある。全てのオブジェクト及び背景は鳥瞰ウインドウのときと同様にアイコンエディタにより作成されている。

【0068】前記における背景は無限遠方にあるという設定なのでユーザが方向を変えない限りこれは変化しないが、ユーザはこの背景と表示されているオブジェクトを頼りに音場空間内を動きまわることができる。

【0069】この実施例におけるデータ表示ウインドウは、ユーザがソースに近づいたときのみ開かれるウインドウである。即ちユーザは任意に(コントロールウインドウのボタンクリックにより)このウインドウを開くことはできない。このウインドウには、ソースの持つ画像情報とテキスト情報が表示される。例えばこの発明を昆虫図鑑に応用した場合には、データ表示ウインドウには図11のように、例えばスズムシの画像とそのデータが表示される。

【0070】前記実施例において、ISFの起動プログラムを実行すると、次の各データファイルを読み込んで

各種の設定を行うことができる。

【0071】(1) ソースデータファイルは、ソースの位置情報、ソースそのものの音量、音の指向性のデータファイルである。

(2) マップデータファイルは、鳥瞰ウインドウ、3Dウインドウに用いるオブジェクトの位置データファイルである。

(3) アイコンデータファイルは鳥瞰ウインドウ、3Dウインドウに用いるオブジェクト、背景アイコンデータファイルである。

(4) 画像・テキストデータファイルは、データ表示ウインドウに表示する画像情報、テキスト情報である。

【0072】前記各データファイルを書き換えることによって、様々なアプリケーションに対応することができる。

【0073】

【発明の効果】即ちこの発明によれば、仮想的な音場空間と、当該音場空間の多数の音源に対応する視覚手段とを具えたインターフェースにより、マウスなどのデバイスを操作して求める音源を検索できるようにしたので、文字データ・数値データその他検索に必要な入力することなく、音と画像をたよりに求める音源を検索できる効果がある。従って求める音源についてのデータが不確かであっても、音を聞き分け音源に到達することによって、正確なデータを表示できる効果がある。

【0074】またインターフェースを多階層化することによって、著しく多量の情報を合理的に整理し、同様の検索により求める情報を得ることができる。

【図面の簡単な説明】

【図1】この発明のISFシステム構成の図。

【図2】同じくハードウェアシステム構成の図。

【図3】同じくMIDIデータフォーマットの図。

【図4】同じくコントロールウインドウの図。

【図5】同じくマウスインターフェースの図。

【図6】同じくレーダウインドウの図。

【図7】同じく鳥瞰ウインドウの図。

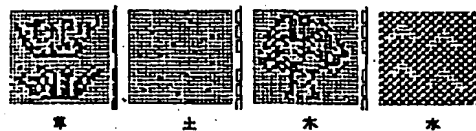
【図8】同じくキャンパスに描かれた鳥瞰ウインドウの図。

【図9】同じく鳥瞰ウインドウ描画のためのアイコンの図。

【図10】同じく3Dウインドウの図。

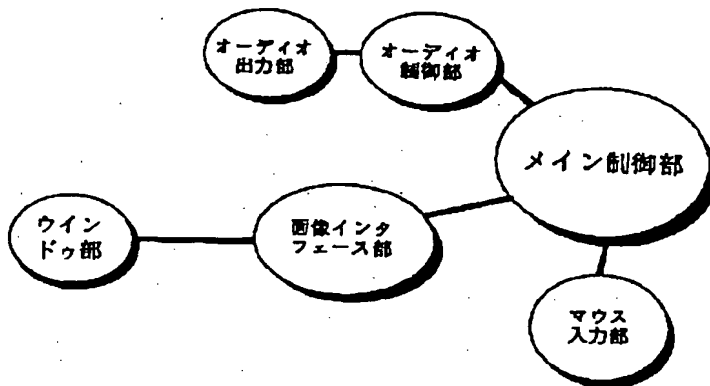
【図11】同じくデータ表示ウインドウの図。

【図9】

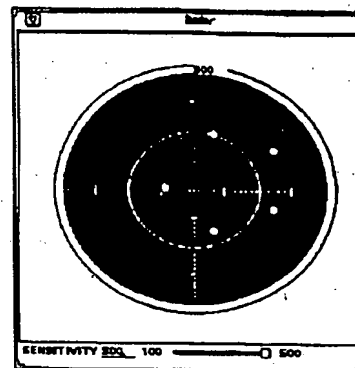




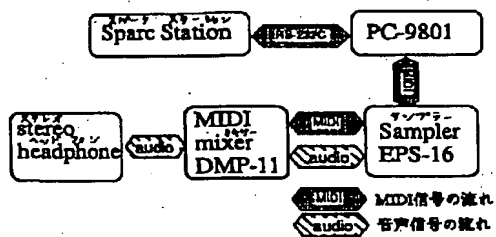
【図1】



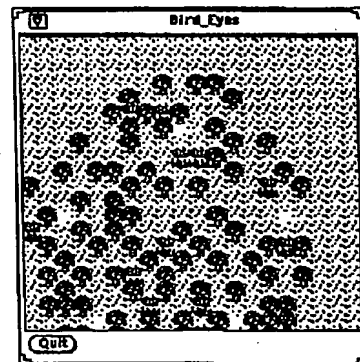
【図6】



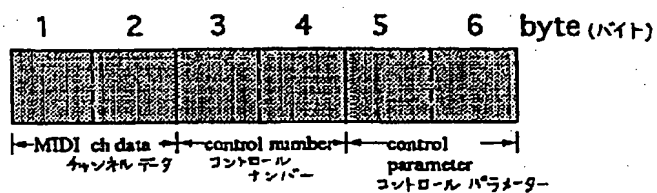
【図2】



【図7】



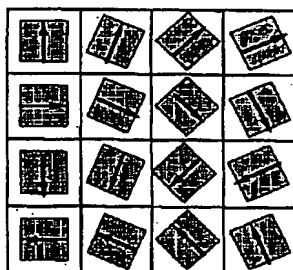
【図3】



【図11】

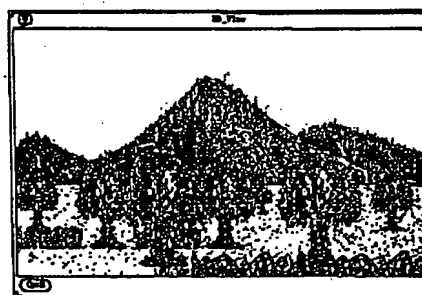


【図8】

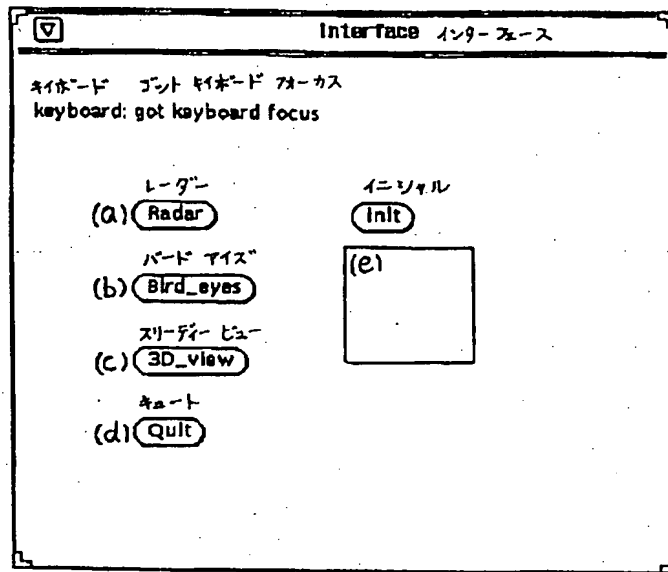


← ユーザーの指示方向

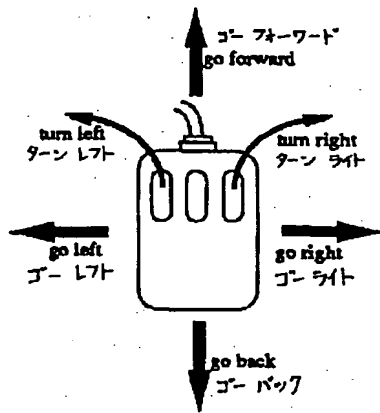
【図10】



【図4】



【図5】



PF030127 PPK  
US

Japanese Kokai Patent Application No. Hei 7[1995]-44575

Job No.: 228-107263

Ref.: JP 7-44575/PF030127 US/PPK (DAVIDA)/#6951

Translated from Japanese by the McElroy Translation Company

800-531-9977

customerservice@mcelroytranslation.com

JAPANESE PATENT OFFICE  
PATENT JOURNAL (A)  
KOKAI PATENT APPLICATION NO. HEI 7[1995]-44575

Int. Cl. <sup>6</sup> :	G 06 F 17/30 G 09 F 27/00 G 11 B 27/10 27/34 G 06 F 15/40
Sequence Nos. for Office Use:	7323-5G 8224-5D 9194-5L
Filing No.:	Hei 5[1993]-192447
Filing Date:	August 3, 1993
Publication Date:	February 14, 1995
No. of Claims:	5 (Total of 9 pages; OL)
Examination Request:	Not filed

SOUND INFORMATION RETRIEVAL SYSTEM AND DEVICE

Inventors:	Atsushi Matsushita 36 Kikui-cho, Shinjuku-ku, Tokyo  Kenichi Okada 4-25-12 Hongo, Bunkyo-ku, Tokyo  Naoto Oki 3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa-ken
Applicants:	391023987 Atsushi Matsushita 36 Kikui-cho, Shinjuku-ku, Tokyo  392008231 Kenichi Okada 4-25-12 Hongo, Bunkyo-ku, Tokyo

Agent:

Masatsugu Suzuki, patent attorney

[There are no amendments to this patent.]

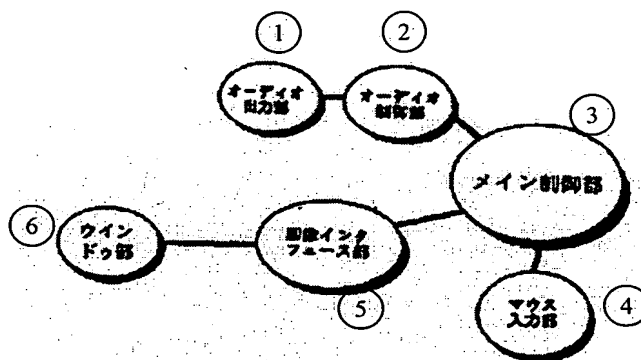
### Abstract

#### Purpose

The purpose of the present invention is to provide a sound information retrieval system, which contains a retrieval interface equipped with a virtual sound field space and a corresponding visualizing means by which means the desired sound source is retrieved from a plurality of sound sources and the data display is realized.

#### Constitution

In the sound information retrieval system, the desired sound source is accessed from a plurality of sound sources on the basis of the direction and distance of the desired sound source. By means of a retrieval interface equipped with a virtual sound field space and a visualizing means corresponding to the plurality of sound sources of the sound field space. Then, by moving in an image and sound field space, by means of a mouse or the like, the visualizing means appears at each prescribed point. When the desired sound source is reached, the image of the sound source and its characteristics, etc. are output. The mouse inputting part which inputs the instructions from the user, an image interface part that stores and provides images, and an audio control part which controls pronunciation are connected to a main controller which controls the entire system, and an audio output part is connected to said audio control part, and an image output part is connected to the image interface part.



Key: 1 Audio output part  
 2 Audio control part  
 3 Main controller

- 4      Mouse inputting part
- 5      Image interface part
- 6      Window part

### Claims

1. A sound information retrieval system characterized in that: by means of a retrieval interface equipped with a virtual sound field space and a visualizing means corresponding to the plurality of sound sources of the sound field space, the desired sound source is accessed from said plurality of sound sources on the basis of the direction and distance of the desired sound source; in that moving in an image and sound field space by means of a mouse or the like, the visualizing means appears at each prescribed point; and in that when the desired sound source is reached, the image of the sound source and its characteristics, etc. are output.

2. The sound information retrieval system described in Claim 1 characterized in that the virtual sound field space is a space for arranging sound sources within the human audible range.

3. The sound information retrieval system described in Claim 1 characterized in that the virtual sound field space is arranged in a hierarchy, and that the accessed regions can be narrowed down.

4. The sound information retrieval system described in Claim 1 characterized in that the visualizing means includes a radar window, a bird's-eye view window, a 3D window, and a data display window.

5. A sound information retrieval device characterized in that a mouse inputting part which inputs instructions from the user, an image interface part that stores and provides images, and an audio control part which controls the pronunciation are connected to a main controller which controls the entire system; in that an audio output part is connected to said audio control part; and in that an image output part is connected to the image interface part.

### Detailed explanation of the invention

[0001]

Industrial application field

The present invention pertains to a sound information retrieval system and device characterized by the fact that it retrieves the desired sound source from a plurality of sound sources by means of a retrieval interface equipped with a virtual sound field space and the corresponding visualizing means and its data to be displayed is received.

[0002]

Prior art

The so-called linking method is the simplest and most often used method for handling expanded functions of multimedia data where the database is mainly for alphanumeric data. In this method, the multimedia data are linked to one or more alphanumeric data, and when said data are retrieved under certain retrieval conditions, the linked image data and sound data are also provided to the user.

[0003]

On the other hand, there is the index method. According to this method, the index retrieval in the conventional database is directly adopted in the multimedia information as is. The file no. or data name of each piece of data and its address in the storage device in which the data are stored are paired to form an indexing table. During retrieval, the user inputs the file number or data name into the retrieval system, and the system checks the input file no. or data name with the index to obtain the address of the data in the storage device, and it then outputs the desired data.

[0004]

A typical example of a system that uses said indexing method is a CD. Each CD has track numbers that correspond to the music selections (data). The user reads the table (index) pairing the data and the track numbers to find the track number of the music selection to be played. By inputting this number in the CD player, the desired music can be played. As a result, the CD itself is an excellent sound database.

[0005]

Another method is the keyword method, which is an improvement over the aforementioned index method in terms of ease of use. A plurality of keywords descriptive of certain features of the data are appended to each data and are used in the retrieval process. This method is adopted by many multimedia databases.

[0006]

In addition, there is the pattern-matching method. Due to developments in recognition technology, multimedia information can be input into the computer and the contents recognized and analyzed. As a result, the sound data can be retrieved under the sound retrieval condition. In this retrieval system, because there is no need to go through the stage of conversion between media as would be required in the case of keyword retrieval (conversion from audio data to text

data, or from text data to audio data), it allows more appropriate sound information retrieval. For example, with a Hyperbook of bird pictures, the user can input the retrieval condition by mimicking the song of the bird to be retrieved. The mimicked song of the bird is analyzed, and the sonic characteristics in terms of amplitude structure, pitch structure, frequency component variation structure, etc. are analyzed, and are compared with the sonic characteristics of the birdsong data that have been registered to determine the distance function.

[0007]

Problems to be solved by the invention

The aforementioned linking method appends the multimedia to the alphanumeric data, making it difficult to perform a direct retrieval, which is undesirable.

[0008]

With the index method, on the other hand, if the user does not know the file number or data name of the information to be retrieved, retrieval is very difficult. For example, if the user does not know the corresponding relationship between the music selection title and the track number on a CD, it is difficult to retrieve the desired music selection (whose title is not remembered).

[0009]

With the keyword method, it is impossible to use keywords to produce a complete representation of the information contained in the audio data and image data. For example, it is difficult to formulate in words a complete representation of the music of "Clair de lune" by Debussy or the atmosphere of the painting of "Melancholy and Mystery of a Street" by Chirico. No matter how hard the user tries to describe the music and atmosphere in words, there is no way that the data and picture can be obtained and no way to check whether the retrieval results are appropriate.

[0010]

The pattern-matching method also has a problem associated with inputting the retrieval condition as a sound. For example, although one may try to mimic the song of a bird, it is difficult to represent it accurately.



[0011]

Problems to be solved by the invention

The purpose of the present invention is to solve the aforementioned problems of the prior art by providing an interface that provides a novel retrieval system that uses virtual reality. In the aforementioned novel system, with the interactive sound field interface, the user sends his/her movement to the interface by means of a mouse or other device. As a result, the interface provides the sound field that corresponds to the movement, so that the user gets a sense of himself/herself moving in virtual space. The next operation takes place in accordance with the movement of the sound field. In this way, the user can move around in virtual space while exchanging information with the interface. Because movements can be made in virtual space, the system has the name interface with interactive sound field (hereinafter to be referred to as ISF).

[0012]

In the ISF, the audio data registered in the audio database is set in a virtual sound field, and the user can move in the sound field space with the direction and distance of the sound emitted by the audio data used as a cue, and it is the retrieval interface that can reach the desired information. By accessing certain information, it is possible to display not only the sound information of the data, but also image data and other media information on the display screen.

[0013]

That is, the present invention provides a sound information retrieval system characterized by the following facts: by means of a retrieval interface equipped with a virtual sound field space and a visualizing means corresponding to the plurality of sound sources of the sound field space, the desired sound source is accessed from said plurality of sound sources on the basis of the direction and distance of the desired sound source; by moving in the image and sound field space by means of a mouse or the like the visualizing means appears at each prescribed point; and when the desired sound source is reached, the image of the sound source and its characteristics, etc. are output.

[0014]

Also, the virtual sound field space is a space for setting sound sources within the human audible range, the virtual sound field space is arranged in a hierarchy, and the accessed region can be narrowed down.

[0015]

In addition, the visualizing means includes a radar window, a bird's-eye view window, a 3D window, and a data display window.

[0016]

Also, the present invention provides a sound information retrieval device characterized by the following facts: a mouse inputting part which inputs instructions from the user, an image interface part that stores and provides images, and an audio control part for controlling pronunciation are connected to a main controller which controls the overall system; an audio output part is connected to said audio control part; and an image output part is connected to the image interface part.

[0017]

As the sound information retrieval interface, the ISF has the following characteristic features:

1. There is no need to input the retrieval condition.
  - (1) No clear retrieval conditions are required.
  - (2) Media conversion of the retrieval conditions are not required.
  - (3) Information retrieval can be performed by browsing.
  - (4) It is a compact interface.
2. The sound image has left/right sensation of position and sensation of distance.
3. A plurality of sound data can be heard simultaneously.
4. It is possible to produce the data space corresponding to the application.
5. It is possible to form a hierarchical structure of the sound information.

[0018]

In the following, a detailed explanation will be given regarding said characteristic features.

[0019]

As explained above, there is the advantage that information can be obtained even when the retrieval condition is not input. This results in several advantages. For example, because there is no need to have a retrieval condition, even when there is no retrieval condition such as the database retrieval in the prior art, retrieval still can be performed if there is an image that asserts "this is actually the sound."

[0020]

Also, because sound is heard as the information is retrieved, it is possible to retrieve the sound information without converting the retrieval condition to another multimedia, such as the keyword retrieval.

[0021]

It is well known in the field of image interfacing that browsing refers to a retrieval method in which the desired information is retrieved while "searching" through information in the same way as a reader browses a book to find certain information. In this case, one may come across useful information while browsing. As a matter of fact, all of us have had the experience, while searching for specific information, of accidentally finding other useful information. Such accidental discoveries of information would be impossible with the conventional database retrieval operation. Browsing is effective as a human interface and can also be effectively applied to sound information.

[0022]

In the conventional database system, the interface for retrieval is very complicated. In many cases, the operation for inputting the retrieval conditions is complicated. Setting the retrieval conditions always requires the character input operation, and a system has been proposed in which all of the operations be performed using a mouse. However, it still requires the operation of setting the retrieval conditions. At this point, with the ISF, it is only required that the user instructs the system by using an input device to indicate where the user wants to move in the sound space. At present, the mouse is used as the input device. However, any scheme may be adopted as long as the direction can be indicated. For example, one may use the number keys on the keyboard or a data glove for the operation.

[0023]

The sound interface of ISF also provides the left/right sensation of position and near/far sensation of distance. Such a sound field with a sense of stereo can realize an environment in which the user can hear the sound data in the natural form. As a result, the user can listen and discern a plurality of sound data emitted at the same time.

[0024]

Humans have the ability to pick out a specific sound in an environment in which there exists a plurality of sounds. For example, suppose that someone is at a cocktail party where many persons are talking loudly; if someone suddenly says that person's name, the person would

immediately turn his/her head towards the speaker. This is called the "cocktail party effect." The cocktail party effect is particularly pronounced when the individual sound sources have sound images with unique left/right position and distance positioning. In this case, by means of reproduction of the individual left/right position and distance positions for the several data registered in the sound database, it is possible to effect identification and comparison.

[0025]

When sound data are arranged in a virtual sound space, because the data have no order, the user will become confused. In this case, one may consider the case when performance is made to mimic certain sounds in the virtual sound space.

[0026]

In ISF, various performances can be made. As an example of the present invention, suppose the system mimics nature in "ISF insect pictures." This space contains meadows, forests, and flowing rivers. In the virtual space composed of their sounds, the sources (sound data) are arranged to mimic "insects in the meadow" and "birds in the forest." Consequently, in the virtual sound space in the ISF, the source (sound data = insects, etc.) and objects (things other than the sound data, including those that do not make sounds = trees, rivers, etc.). The user can walk in the virtual sound space to get the information.

[0027]

Although humans have the ability to exercise the selective hearing of the cocktail party effect, it is impossible to hear and discriminate hundreds and thousands of sounds at the same time. On the other hand, there are also hardware restrictions. In the test system assembled for this test, the limit is that up to 8 types of sounds can be sent independently to the audio line. Consequently, from the human side, the types of sounds that can be generated simultaneously are limited. In this case, in order to realize practical application, it is preferred that a hierarchical structure of the sound information be formed.

[0028]

For example, the sound information may be divided into several groups beforehand, and the sound information belonging to a certain group is positioned in a range up to a certain position in the virtual sound space. When the user is outside a certain group, the representative sound of the group can be heard as the sound image position at the center of gravity of the group. The representative sound of the group may be a single sound or all of the sounds of the entire group. Only when the user enters the group, can a sound be generated with individual sounds in

the group having respective sound image position. As a result, a plurality of sound information can be handled.

[0029]

For example, among insects, a group of cacadas is formed. When the user approaches from outside the group, the user can hear a chorus of chirps from the cacadas. Then, once the user enters the group, the user can hear the individual cacadas chirping at their respective sites.

[0030]

The method using effector may be adopted in assembling said ISF. In this method, the sensation of distance of the sound can be represented in a virtual way. As a result, in the window, it is possible to display the window and the back-and-forth relationship of the icon as the sensation of distance. For example, in order to represent the remote side, the effect of the reverberation (residual sound) is increased. On the other hand, when the sound from the back side is to be represented, a low-pass filter or the like is used. With this method, in addition to the left/right sound volume difference, the sound image position in the left/right direction can be realized by means of the time difference between the left/right signals.

[0031]

With said effector method, it is possible to exchange information with the computer by means of the interface of the MIDI specifications, and control can be performed easily, which is advantageous.

[0032]

However, when the ISF of the present invention is assembled, in addition to said effector method, one may also adopt the correlation coefficient variation method in which the correlation coefficient between the right signal and the left signal is changed continuously from 1 to -1.

[0033]

In a binaural system, sound is recorded by means of two microphones placed at portions corresponding to the eardrums of the two ears of a dummy head that acoustically mimics the human head. When the user hears the recorded sound over headphones, it is possible to use the binaural system to realize the sensation of position the sound image at the position of the dummy head. In particular, for the binaural system adopted at present, without using the dummy head, the transfer function of the acoustic properties of the dummy head is measured, and the digital filter representing it is used to realize the binaural effect. With this system, it is possible to

realize sound image positioning in the back-and-forth and left/right directions. This is a characteristic feature of this scheme.

[0034]

The aforementioned systems each have their advantages and disadvantages. Consequently, when the invention is embodied, a system that matches the characteristics of the object is adopted. For example, it is preferred that assembly be performed using the effector method in retrieval of insects containing a lot of high frequency components.

[0035]

Also, in the binaural system, the time required to compute the convolution integral performed using digital filter becomes very long, and a real-time reaction for the overall system cannot be realized. If this problem can be solved, the method can well be adopted.

[0036]

In summary, there are no restrictions on the assembly system of the present invention. One may adopt the appropriate system for the given object sound source.

[0037]

Operation

According to the present invention, because the desired sound is focused on from among a plurality of sounds, there is no need to input the retrieval value and other symbols beforehand.

[0038]

Also, one may adopt a scheme in which the virtual sound field space and a visualizing means are combined to improve the user's on-site sensation, so as to facilitate his making an appropriate selection of the desired sound.

[0039]

Then, by using a plurality of visualizing means at the same time, it is possible to move towards the desired sound source, so as to narrow down the desired sound source.

[0040]

Application examples

In the following, an explanation will be given regarding an application example of the present invention with reference to Figures 1 and 2.

[0041]

Figure 1 is a diagram illustrating the constitution of the hardware in the present invention. In this constitution, the main controller is connected to the mouse inputting part, the image interface part, and the audio control part. The audio control part is connected to the audio output part, and the image output part is connected to the image interface part.

[0042]

Figure 2 is a diagram illustrating the equipment on the basis of the constitution shown in Figure 1. In this case, a sparc station is connected via RS-232C to personal computer NEC PC-9801; PC-9801 is connected via MIDI to sampler EPS-16; sampler EPS-16 is connected via MIDI, Audio to MIDI mixer DMP-11; and MIDI mixer DMP-11 is connected via Audio to stereo headphones.

[0043]

Said sparc station (1) is a device used to control the entire system. Here, by means of the display, the visual information is provided, and the mouse is used to move around in the sound field space.

[0044]

Said DP/4 can independently process each of the independent audio signals of the 4 channels. Such processing can be performed in real time by means of MIDI signals (Figure 2). Because the correlation coefficient variation method is used, phase control of the audio signal is performed for control of the sensation of distance of the sound image.

[0045]

Besides the audio signal processing described above, it is possible to provide phase shifter callers for reverberation (reflected sound effect) and delays for realizing the sensation of distance and broadening of the sound image, and it is possible to provide equalizer functions for changing the sound quality, etc.

[0046]

Also, EPS-16 is the sound source of the source object, and it can be controlled by the MIDI signal. In a 1-MB memory, up to 32 types of sounds can be digitally recorded, 16 types of sounds can be reproduced simultaneously, and audio signals can be output from the independent

lines of 8 channels. The sampling rate of said EPS-16 is up to 44.8 MHz, and it is possible to obtain the same sound quality as that of CD.

[0047]

DMP-11 is used to group the plurality of sound images obtained with said EPS-16 and DP/4 to a single sound field. Here, DMP-11 is an 8-channel digital mixer that allows control with MIDI signal. The analog sound lines of 8 channels are A/D-converted when input, and then the unmodified digital signal is sent to a DSP for processing. Consequently, it is possible to suppress the degradation of the sound quality and to suppress generation of noise, and, at the same time, it is possible to realize panning (left/right sound image positioning), reverberation (reflected sound effect), equalization (change in frequency component), etc. independently for each channel (Figure 2).

[0048]

As explained above, it is preferred that said MIDI equipment be used and the workstation support the aforementioned functions. However, in the present development state, only a portion of the function can be provided. For example, instead of a sampler, it is possible to store the audio data in a peripheral storage device of the workstation. However, the audio data cannot be output from two or more channels as a plurality of independent lines. In the future development of multimedia, when said functions are incorporated as a portion of the workstation, they can also be adopted in the present invention.

[0049]

There are the following four parameters pertaining to the sound image positioning in the sound image control in said application example.

[0050]

1. The volume is inversely proportional to the square of the distance between the user and the source object.
2. For left/right positioning, panning is used to divide the sound image into 32 directions spanning 360°.
3. For the sensation of distance, reverberation is used so that the greater the distance, the higher the reverberation level.
4. The sound that should be heard from behind the user is that obtained by means of a low-pass filter.



[0051]

When the correlation coefficient variation method is used to generate the aforementioned sensation, relatively good results can be obtained as far as white noise is concerned. However, in this application example, when the sound of insects are retrieved, a good sensation of distance cannot be obtained. Consequently, in this application example, the method using effector is adopted. As explained above, when the sounds of insects are retrieved, the correlation coefficient variation method gives unfavorable results. The reasons are as follows.

[0052]

That is, the correlation coefficient variation method has the sensation of distance appear significantly with respect to a sound image moving in the back-and-forth direction; the sensation of distance depends on frequency, with a less pronounced sensation of distance in the higher frequency region; when practically adopted in broadcasting, etc., it is combined with the other methods for displaying the sensation of distance, etc.

[0053]

Consequently, depending on the property of the sound source, there is a method that can be well adopted for embodiment of the invention.

[0054]

The MIDI data that control said DMP-11 and EPS-16 are individual 6-byte information [units], and the structure shown in Figure 3 is adopted. That is, the first and second bytes are the channel numbers assigned to the controlling MIDI equipment. For example, the channels are assigned as follows: DP/4 is assigned to ch 1, EPS-16 plus is assigned to ch 2, and DMP-11 is assigned to ch 3 and ch 4. The second and third bytes are the numbers of the functions that should be controlled, and the fourth and fifth bytes are the parameters of the function. Their values are determined on the MIDI equipment side, and they are adopted.

[0055]

In said MIDI data generating part, the various parameters for positioning the sound image computed with the sound image position computing part are converted into MIDI signals in said data format. The MIDI data formed in said MIDI data generating part are output from the RS-232C port of the sparc station. In the current version, a handshake or another error inhibiting procedure is not used. Consequently, the measure for the communication error is adopted in the MIDI signal relay part. In the present invention, in addition to the sound interface, the direction and sensation of distance of the sound source are judged. At the same time, as an auxiliary

interface that represents the sensation, several interfaces using images are adopted. Consequently, an example will be explained. This image interface is realized on the X-window by means of the X-view. Figure 4 is an example of the control window.

[0056]

This control window is the unique window present on the initial picture when the system of the present invention is started. The mouse operation by the user is mainly performed in this window, and the mouse interface portion also belongs to this control window.

[0057]

Buttons (a-c) in this window open windows onto the various images; button (d) ends the system; and canvas (e) is used to sense the movement of the mouse.

[0058]

By moving the mouse in said canvas (e), the user can enter the sound space in this direction. Also, by means of the left/right buttons on the mouse, the user can change the direction stepwise in steps of  $360/16 = 22.5^\circ$ . That is, as shown in Figure 5, the user can move in the sound space by selecting any of the 16 directions.

[0059]

In the mouse interface part, the position of the user is changed by movement, and the information is transmitted to the window and audio control part. Also, in the mouse interface part, control of the position of the mouse is performed in the control window. After the mouse enters canvas (e) (Figure 4), as a result of movement in the sound field space, the mouse comes out of the canvas, and the position of the mouse is controlled to return to the center of the canvas. Of course, even when the mouse returns, there is still no change in the position of the user. That is, the user can get to anywhere in the sound field space regardless of the position of the mouse in the control window. As a result, the user even can close his eyes while moving by relying solely on the sound, or the user can open his eyes to watch another window while moving.

[0060]

In this application example, for the radar window, the source is a window that displays the figure (Figure 6) as seen from above the virtual space, and the object is not displayed. Because the relative position between the source and the user is judged, the mouse can be easily moved while watching the source.

[0061]

As shown in Figure 6, the user is displayed as a white spot on the radar screen, and no distinction is made according to the type of source. Also, the direction corresponding to the front side of the user is the upper direction of the vertical line (upper side on the screen), and there is no change even when the button on the mouse is clicked. For example, when the user clicks the left button of the mouse and makes left rotation, the radar window has the source rotated to the right around the user at the center, and the user is always kept facing forward (upper side of the radar screen). On the radar window, a slider bar (Figure 6b) called "sensitivity" is attached, and it is used to change the radar sensitivity. By changing this sensitivity, the range shown on the radar is changed, and its value is displayed on the side of the slider bar and on the upper side of the radar screen.

[0062]

A large value of sensitivity corresponds to a wider range shown on the radar, and a small value of sensitivity corresponds to a narrower range. The sensitivity can be changed as desired. Consequently, when it is not near the source, the sensitivity becomes higher, and, as it approaches the desired source, the sensitivity becomes smaller, and the position of the user is finely adjusted. Also, the dashed pattern of the horizontal line and vertical line of the radar screen can be changed according to the sensitivity value such that the visual sensitivity can be judged.

[0063]

In this application example, a bird's-eye view window (Figure 7) has the relationship of inner/outer with the radar window. Without displaying the source, only the object is displayed, and it shows the picture as seen from above the scene of the periphery of the user.

[0064]

In this window, the user moves towards the source at the site of the source (such as in the forest for cacadas).

[0065]

In this window, the user is positioned at the center of the screen, as in the radar window, and the upper direction of the screen is in agreement with the front side of the user. The bird's-eye view window becomes the field of view of the user corresponding to the 5 objects on the periphery as well as back-and-forth and left/right sides of the user from among the map

corresponding to, e.g., the 50 x 50 map data. Also, scrolling can be performed by the user moving the mouse, and rotation can be performed by clicking the mouse.

[0066]

In the embodiment, for example, as shown in Figure 8, 16 maps of 50 x 50 are drawn in 16 directions on the canvas, and a bird's-eye view window shows some of them. As the user clicks the button of the mouse to input the information, it jumps to the map corresponding to the direction. As a result, the front side of the user remains unchanged at all times. As a result, although preparation of 16 types of maps is unfavorable in terms of memory efficiency, it nevertheless can lead to reducing the time for reloading. Also, no snow phenomenon takes place for the picture shown in this case. In this method, it is possible to realize real-time operation. The bit map using the aforementioned bird's-eye view is formed from the icon editor, as shown in Figure 9. In practice, the size of the object as seen by the user becomes  $16 \times 16 = 256$  dots. Because the bit map is square, if  $16 \times 16$  is used, there exists no space on the map when the magnification is other than the magnification of the direction of  $90^\circ$  which the user faces. When a  $24 \times 24$  bit map is used, since the objects is used for the superimposed portion are ORed, the presence of said space is excluded.

[0067]

In this application example, the 3D window (the window of the three-dimensional object) (Figure 10) displays only the scene on the periphery of the user, as in the case of the bird's-eye view window. Here, a three-dimensional object is set. It is for setting the scene near the line of sight of the user so as to improve the on-site sensation. Also, in this window, the background is also displayed for direction sensing and recognition of the direction by the user. The background that enters the picture of the window is taken as the range corresponding to  $90^\circ$  of the angle of the field of view of the user, and the object entering this field of view is displayed on the screen. Because there are 16 directions that the user can take, 16 backgrounds are prepared, and by clicking the mouse once, the background is changed for a quarter of the screen in each step. According to the distance to the user, 4-6 types of objects are prepared. All of the objects and the backgrounds are prepared by means of the icon editor in the same way as in the case of the bird's-eye view window.

[0068]

Because said background is set at an infinite distance, it does not change as long as the direction of the user is not changed. However, the user can move in this sound field space by relying on the background and the displayed object.

[0069]

In this application example, the data display window is the window that is opened only when the user approaches the source. That is, the user cannot open the window at will (by clicking the button on the control window). In this window, the image information and text information of the source are displayed. For example, when the present invention is adopted for the insect picture collection, as shown in Figure 11, for example, the image of a "bell-ring" insect and its data are displayed in the data display window.

[0070]

In said application example, as the start program of ISF is executed, the various data files are next read to make various settings.

[0071]

(1) The source data file is a data file of the position information of the source, the volume of the source itself, and the directionality of the sound.

(2) The map data file is a data file of the positions of objects used in the bird's-eye view window and the 3D window.

(3) The icon data file is a data file of the objects and background icons used in the bird's-eye view window and 3D window.

(4) The image/text data file includes the image information and text information displayed on the data display window.

[0072]

By means of rewriting said data files, it is possible to perform various applications.

[0073]

#### Effect of the invention

According to the present invention, by means of an interface having a virtual sound field space and a visualizing means corresponding to the plurality of sound sources in the sound field space, it is possible to retrieve the sound source determined by manipulating the mouse or other device. Consequently, there is no need to input the alphanumeric data that used to be necessary for retrieval, and the sound source can be retrieved by means of sounds and images. Consequently, even when the data for the desired sound source is not confirmed, it is still possible display the correct data by hearing the sound to reach the sound source.

[0074]

Also, by adopting a hierarchical structure for the interface, information in much larger quantities can be arranged reasonably, and it is possible to obtain the information determined by means of the same retrieval.

#### Brief explanation of figures

Figure 1 is a diagram illustrating the ISF system constitution of the invention.

Figure 2 is a diagram illustrating the constitution of the hardware system of said ISF system.

Figure 3 is a diagram illustrating the MIDI data format of said ISF system.

Figure 4 is a diagram illustrating the control window of said system.

Figure 5 is a diagram illustrating the mouse interface of said system.

Figure 6 is a diagram illustrating the radar window of said system.

Figure 7 is a diagram illustrating the bird's-eye view window of said system.

Figure 8 is a diagram illustrating the bird's-eye view window depicted on the canvas of said system.

Figure 9 is a diagram illustrating the icon for depicting the bird's-eye view of said system.

Figure 10 is a diagram illustrating the 3D window of said system.

Figure 11 is a diagram illustrating the data display window of said system.

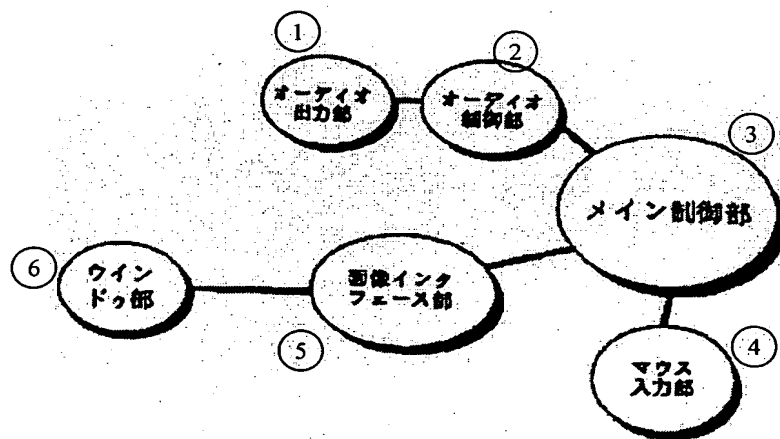


Figure 1

- Key:
- |   |                      |
|---|----------------------|
| 1 | Audio output part    |
| 2 | Audio control part   |
| 3 | Main controller      |
| 4 | Mouse inputting part |
| 5 | Image interface part |

## 6 Window part

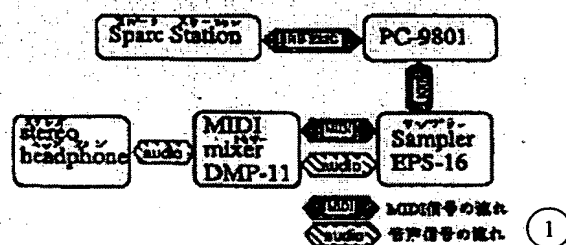


Figure 2

Key: 1      Flow of MIDI signal  
              Flow of audio signal

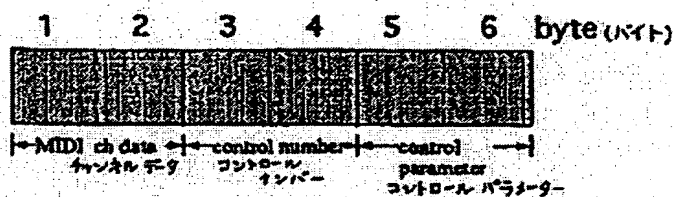


Figure 3

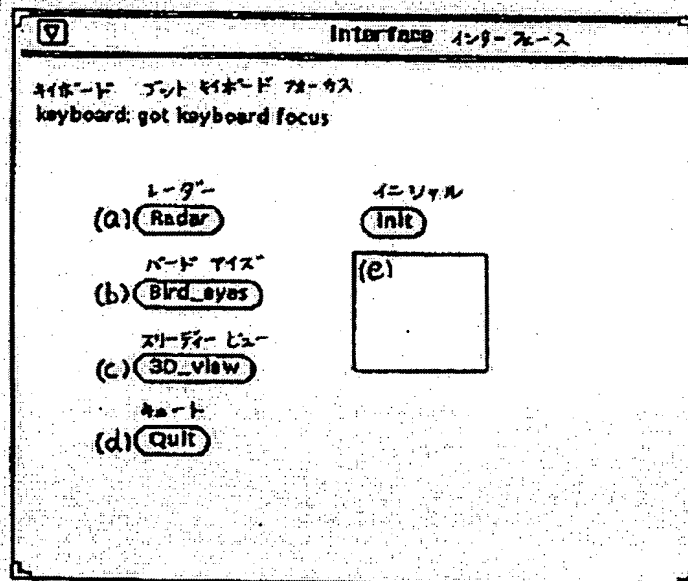


Figure 4

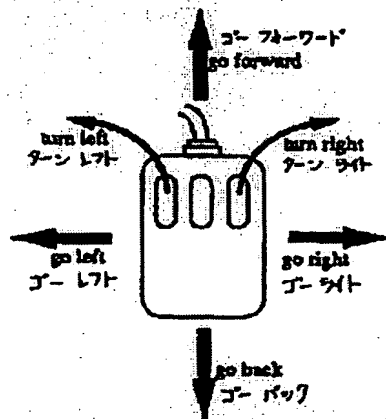


Figure 5

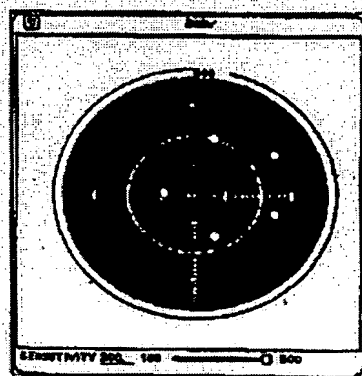


Figure 6

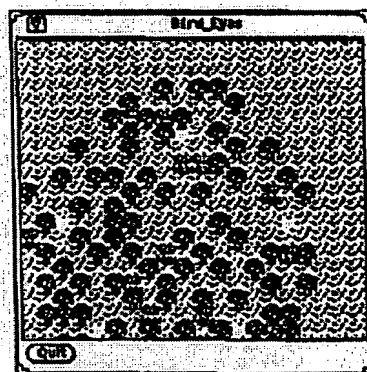


Figure 7



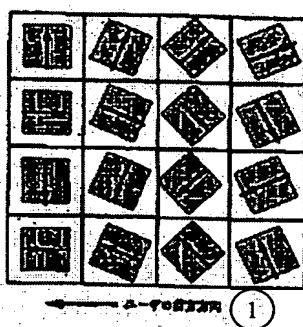


Figure 8

Key: 1 Front direction of user

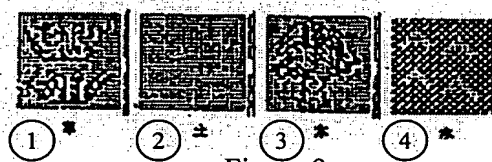


Figure 9

Key: 1 Grass  
2 Soil  
3 Trees  
4 Water

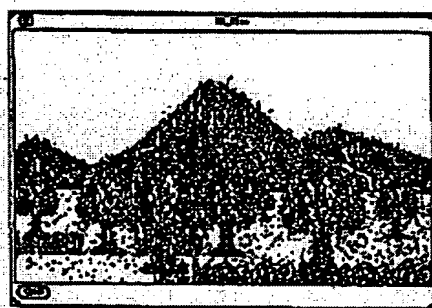


Figure 10



Figure 11

Key: 1 [illegible]

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**